Product Recommendation System for Supermarket

Pranavi Satheesan
Department of Information Technology
Srilanka Institute of Information Technology
Colombo, Sri Lanka
it17095136@my.sliit.lk

Prasanna S. Haddela
Department of Information Technology
Srilanka Institute of Information Technology
Colombo, Sri Lanka
prasanna.s@sliit.lk

Jesuthasan Alosius
Department of Information Technology
Srilanka Institute of Information Technology
Colombo, Sri Lanka
jesuthasan.a@sliit.lk

2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00151

Abstract— **Customers who seek the services at supermarkets are subjected to inconsistencies & ambiguities over choosing their desired products from a wide range of products with the closest quality. Meanwhile, supermarkets find it very difficult to satiate the customers' demand. Therefore, proposing a method to analyze the customers' need plays an important role in attracting new and regular customers. The purpose of this study is to formulate a product recommendation system which analyze customers' needs and thus recommend the best products. This system recommends products to the regular customers and to the new customers as well. New customers mean obviously the customers with no purchasing history at the supermarket in question. The system referred to recommends the products to the new customers using up two methods. One method recommends the most popular products while the other method solely focuses on the product description for recommendation. The system recommends the products to the regular customers using up user-based collaborative filtering, item based collaborative filtering and association rule mining. It recommends products to regular customers based on purchasing history and priority ratings given by other users who bought the products. Initially, the recommendation algorithm finds a set of customers who purchased and rated the products that overlap with the user who purchased and rated the products. The algorithm aggregates products from the customers with similar preference and eliminates the products the user has already purchased or rated. The proposed methodology improves the shopping experience of customers by recommending accurately and efficiently the products that are personalized to the need of the customers.**

KEYWORDS: recommendation system, collaborative filtering, cosine similarity, association rule mining, correlation

I. INTRODUCTION

Recommendation systems play an important role in supermarkets. Recommender Systems are intelligent engines that gather information relevant to which product a customer has seen or bought previously, with the objective of providing personalized suggestions on unobserved items that are likely to be interest [1]. It analyses the needs of customers and suggest the best possible shopping list. Most of the customers would like to have recommendation system because they can gain access to know about the feedback given by other customers. Many applications use the products that customer purchased and explicitly rated to reflect their interests. There are three popular algorithms used in the recommendation systems. They are collaborative filtering, cluster models and association rule mining. Collaborative filtering algorithms recommend products based on the opinions of other like-minded customers. There are two types of collaborative filtering. They are item-based collaborative filtering and user-based collaborative filtering. In item-based collaborative filtering and user-based collaborative filtering, similarities between items and users are calculated using cosine similarities respectively. User-based method uses historical information to identify the neighborhood for the active customer. In consequent to this, the products are recommended according to their similarities to this neighborhood. User-based recommender systems use the customer profile data and they can incorporate demographics of customers along with the purchasing data history [2]. Recommendations for customers are computed by finding products that are similar to other products preferred by the customers. Cluster methods are used either in case where a supermarket system is to recommend products to a new customer or in a new supermarket that is obviously lacking purchase history of customers. This recommendation is based on textual clustering analysis given in product description.

Nowadays with the using up of data mining and applying the rules in this field, we can create models on data that reveal this implicit knowledge and pass on information to us [3]. We use data mining in product recommendation system because the customers can have easy access to their desired products without consuming much time for search. Monitoring commercial transactions can mine patterns as association rules to discover the potential relationship between the products in the store and subsequently suggest the products with similar quality to customers. In addition, the product recommendation systems can monitor the history of purchasing behavior of customers, their preferences and predict the needs of customers and finally propose the closely related products to customers. Customers can also recognize products based on the recommendations that associate with their priorities, and make a final decision for the purchase. Hence, the recommendation system can be helpful for the users to identify suitable products for their needs and preferences in an effective manner.

Rest of the sections of the paper has organized as follows. In research problem section addresses the difficulties the customers are facing in supermarkets. Related work section summarizes the relevant work in the chosen area. Proposed method section explains procedures and techniques used. Results & Discussion sections explains the results and limitations of the project. The conclusion section explains the aims and objectives achieved and also the improvements which can be done in the future.

II. RESEARCH PROBLEM

It is a difficult task to find the personal preference of customers on a wide range of choices offered for a particular product. It is unable to make sure whether the customers tend to purchase accessories with their favorite main products. Customer preferences will always change beyond your imagination. Customers are subjected to difficulties at supermarkets in selecting products from a large variety of

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00151

products. They spend a lot of time in finding their desired products. Majority of people didn't plan about the contents of their purchase before the inception of supermarkets. They are unable to remember the list of goods while purchasing. Such condition prevails when people attempt to buy things for a social event/official ceremony or to prepare certain cuisines. We all know that the customers purchase some particular items on daily, weekly, monthly basis in compliance to the varying up seasonal requirement. In case where a new customer enters into the supermarket, he /she finds it difficult to choose products due to the lack of purchasing history that instils confidence and recommendation which can be simply interpreted as cold start problem.

III. Related Work

The study [4] recommends the items to be added to the basket. Two approaches are proposed here. First one explores a factorization-based model called Basket sensitive Factorization Machine (BFM) which incorporates different types of associations involving the user, the target item to be recommended, and the items currently in the basket. Second one is based on the observation of various recommendations towards constructing the same basket with similar likelihoods, proposing another model called Constrained Basket sensitive Factorization Machine (CBFM) that further incorporates basket-level constraints.

The research paper [5] suggests products to a new user. Recommendation of the products depend on the buying pattern of previous users whose buying pattern are close to that of new user. To find out the nearest user profile among the profiles of all users, the system uses vector space model and Single Cycle Multiple Attractor Cellular Automata (SMACA) in implementing association rule.

The article [6] proposes prediction model based on the behavior of each customer using data mining techniques. The model analyses the data in order to classify both customers and products. The model classifies customers according to their buying pattern and recommends new products more likely to be purchased by them. The prediction model is intended to be utilized as a tool for marketeers to provide a targeted and specified consumer behavior.

The [7] K-Apriori algorithm is used to split the customers into different categories. Then, frequent item sets and association rules are found those categories. The K-Apriori algorithm tries to find the purchase patterns as groups and those specific groups of people can be satisfied effectively. Related products are placed together in such a manner to the customers to increases the customer satisfaction and the customers can find items logically they tend to buy among the rest.

The research paper [8] proposes a system that recommends products to a new user. Recommendation depends on the buying pattern of previous users whose buying pattern is closed to that user who asks for a recommendation. To find out the nearest user profile among the profiles of all users in database, the system used weighted cosine similarity measure The Association rule mining rule is also used in recommending the products.

In the article [9] recommendation system is developed for e-commerce sites using collaborative filtering. The proposed method analyses the behavioral patterns of the customers, finds the preference levels of a customer for the products which are clicked but not bought, and the collaborative filtering is conducted using the preference levels for providing recommendations.

The research paper [10] introduces an approach that combines entropy-based algorithm, clustering and Bayesian interference in order to minimize cold-start problem, data sparsity and other scalability issues. The approach captures the dynamic environment of a supermarket as users change their preferences time to time. Traditional collaborative filtering technique is used to achieve this.

Most of the previous researches did not recommend products to the new customers. It only considers recommending products to the existing customers. This research focuses on providing recommendations to both existing customers as well as new customers.

IV. Proposed Method

Recommendation system is built by the techniques namely cosine similarity, collaborative filtering and association rule mining.

A. Cosine Similarity

It is the measure of similarity between two vectors by calculating the cosine of the angle between them. The algorithm generates recommendations based on customers who are very similar to the user. It can measure the similarity between two customers and two products by measuring the cosine of the angle between two vectors.

B. Association Rule Mining

Association rule is mainly used to identify the interesting hidden relationships among attributes of huge data set. Association rule mining finds repeated items in a set of transactions as frequent patterns. In this case, Apriori algorithm is used to discover relationship among products.

C. Collaborative filtering

Recommends products to customers based on purchase history and similarity of ratings provided by other users who bought items to that of a particular customer.

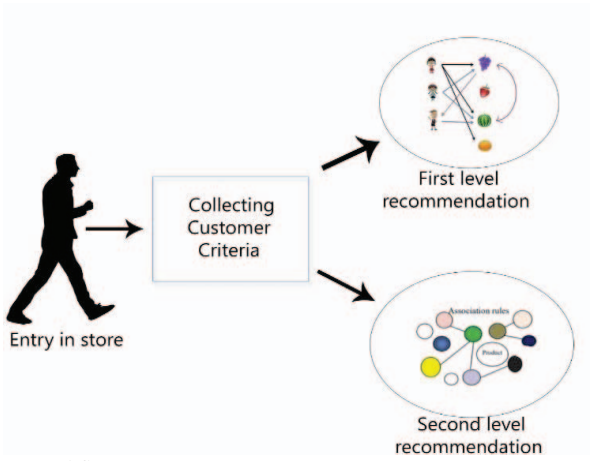

Figure:1:System Diagram

Figure 1 shows the system diagram that includes the product recommendation at two stages. In the first stage, recommendation is done in prior to the purchasing of product. In the second stage, recommendation is done in consequent with the purchasing of product. Firstly, data on previous purchases of products are collected. Initially, product recommendation system recommends the products that have closely related with the criteria of customers.

User-based systems use the customer profile data. They can incorporate demographics data along with the historical purchasing data to identify the closest customer for the active customer. Products are recommended by using their similarities to this neighborhood. Recommendations are based on calculating the similarities of two users. User-to-user matrix is built by iterating through all user pairs and computing similarity matrix for each pair. Similarity between two items is measured using up cosine similarity. Similar products at the top are recommended based on cosine similarity. Giving recommendation to the user B is depending on the buying pattern of the user A.

In contrast, the item-based method uses only the purchasing data history to identify similarities between different items. In the Item-based method, a table with similar items is built by finding the items on which the customers have a tendency to buy together. Therefore, product-to-product matrix is built by iterating through all item pairs and computing similar matrix for each pair. Similarity between two items is measured using up cosine similarity. The algorithm finds similarity in each purchase and the ratings of the user, aggregates the items purchased and recommend the most popular items.

In the second stage, it recommends the purchasing of associated products with the desired product of customers to complete the buying process and to make customers aware of potentially related products with their desired products. The information about the history of shopping behavior includes the products that are purchased very often with other products. As a result, the relationship between products can be explored in terms of a data mining application called association rule mining. Using these rules, we can find buying patterns. The relationship between the products will steadily increase the likelihood for buying the desired products with associated products.

New customers mean obviously the customers with no purchasing history. The system recommends products in two ways. In one method, the most popular products are recommended to the new customers. Most popular products are identified through ratings given by the regular customers of the supermarket. In the other method, products are recommended based on product description. K-means clustering is used to find top words in each cluster on the basis of product description. In case where a word appears in multiple clusters, the algorithm chooses the cluster with the highest frequency of occurrences of the word. The recommendation system display items from the corresponding product clusters based on the product descriptions.

V. Results & Discussion

The proposed system improves shopping experience of the customer. When a new customer visits the supermarket for the first time, the system recommends the most popular products. Most popular products are identified by the ratings given by the regular customers of the supermarket. The following figure 2 shows the pattern of recommendation of top 5 popular products to new customers based on ratings.

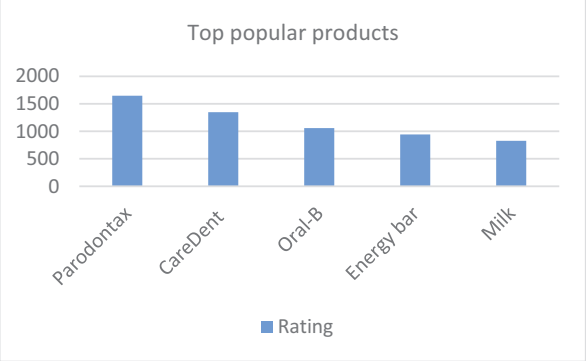

Figure 2: Top 5 recommended products

In case where a new customer searches for some specific product, the product recommendation system recommends products based on textual clustering analysis given in the product description. K-means clustering is used to find top words in each cluster based on product description. If a word appears in multiple clusters, the algorithm chooses the cluster with the highest frequency of occurrences of the word. The system display items from the corresponding product clusters based on product descriptions. Figure 3 shows that the products are clustered into 5 groups. Each cluster contains the top 5 products.

```
Top products per cluster:
Cluster 0:
 Elephant house icecream strawberry 500ml
 Highland processed cheese
 Snack cracker
 body spray
 Promate Exercise Book single rule 280 pages
Cluster 1:
 NSP Exercise Book single rule 360 pages
 SchoolMate Exercise Book single rule 240 pages
 Promate Exercise Book single rule 280 pages
 spaghetti
 SchoolMate Exercise Book square rule 120 pages
Cluster 2:
 Kotmale fresh cream croissant
 Graph Book 160 pages
 Jason sea fresh
 Rathna Exercise Book single rule 400 pages
 Highland processed cheese
Cluster 3:
 Richard Exercise Book single ruled 120 pages
 Highland Ghee
 Highland Ghee
 Elephant house icecream faluda 400ml
 Dilmah Premius Ceylon Tea 500g
Cluster 4:
 Dilmah Premius Ceylon Tea 2000g
 Dr Pepper tin soda
 Atlas Exercise Book single ruled 320 pages
 NSP Exercise Book single rule 280 pages
 Ginger soda 1250ml
```

Figure 3: Top products per each cluster

Figure 4 shows that in case if a customer searches “blue bowl” it first identifies the best cluster and here it is found as cluster 0. Then, the system displays the products which is clustered into cluster 0.

```
show_recommendations("blue bowl")

Cluster 0:
 Elephant house icecream strawberry
 Highland processed cheese
 Snack cracker
 body spray
 Promate Exercise Book single rule
```

Figure 4: Recommendation using clusters

The system recommends products to the regular customers using up two techniques such as user-based collaborative filtering and item-based collaborative filtering. The User-item rating matrix is built at first in both techniques. Table 1 shows the user-item matrix which is a 5*5 matrix. But this matrix is built for all users along with all products in the dataset.

TABLE 1:USER-ITEM MATRIX

ProductID / CustomerID	10002	10080	10135	15030	21874
1069	0	0	0	0	0
5517	0	0	1	0	0
6668	0	0	0	0	0
12249	0	1	0	0	0
13697	0	0	0	0	1

Similarity between two users is calculated in user-based technique. Similarity between two users is calculated using up cosine similarity. The algorithm builds a similar-user table to determine the most-similar match for a given item. Table 2 shows the user-user matrix which is a 5*5 matrix. The User-to-user matrix is built by iterating through all user pairs and computing a similar matrix for each pair.

TABLE 2:USER-USER MATRIX

CustomerID / CustomerID	22063	22066	22070	22071	22072
22063	1.000	0.000	0.000	0.258	0.143
22066	0.000	1.000	0.000	0.000	0.064
22070	0.000	0.000	1.000	0.000	0.078
22071	0.258	0.000	0.000	1.000	0.166
22072	0.143	0.064	0.078	0.166	1.000

Similarity between two items is calculated in item-based technique. Similarity between two items is calculated using up cosine similarity. The algorithm builds a similar-item table to determine the most-similar match for a given item. Table 3 shows the item-item matrix which is a 5*5 matrix. The Item-to-item matrix is built by iterating through all item pairs and computing a similar matrix for each pair.

TABLE 3:: ITEM-ITEM MATRIX

ProductID / ProductID	10002	10133	10135	21876	21877
10002	1.000	0.008	0.008	0.025	0.012
10080	0.008	1.000	0.010	0.019	0.007
10120	0.008	0.010	1.000	0.028	0.035
10125	0.025	0.019	0.028	1.000	0.036
10133	0.012	0.007	0.035	0.036	1.000

With the use of similar-items table, the algorithm finds items similar to each purchase and rating of users, aggregates those items and then recommends the most similar items. This computation is very quick and depending only on number of items purchased or rated by the user. Figure 5 shows the recommended products to the regular customers of supermarket.

Product_Name
Almonds
whole wheat rice
Signal
Blueberry jelly
captain fish
closis sensitive flouride
gillette vector
Listerine Essential care
Milk Shorties
Elephant house icecream berry 450ml

Figure 5: Recommended List

A survey is conducted for among the people who are using supermarkets for purchasing products in order to identify what the people think about the product recommendation system.

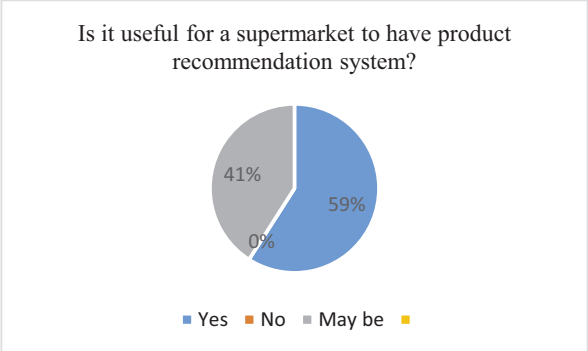

Figure:6: Pie-chart of usefulness of product recommendation system

Figure 6 shows that 59.1% of the people felt that it is useful to have a product recommendation system in the supermarket and 40.9% felt that product recommendation system may be useful for supermarkets. No one said that it is not useful to have a product recommendation system. Most of the people

said that it is useful therefore, it clearly states that the product recommendation system is useful in supermarket.

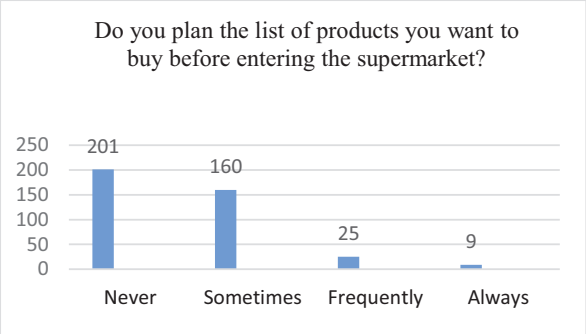

Figure 7: Bar chart of whether people plan before going to supermarket

Figure 7 shows that out 386 people, 173 people sometimes plan the list of products before entering the supermarket, 153 people never plan the list of products before entering the supermarket, 51 people frequently plan the list of products before entering the supermarket and the rest 9 people always plan the list of products before entering the supermarket. According to the results as most of the people sometimes plan the list of products before entering the supermarket it states that they don't have clear idea about what they are going to buy. Therefore, the product recommendation system can able to help the people in purchasing products.

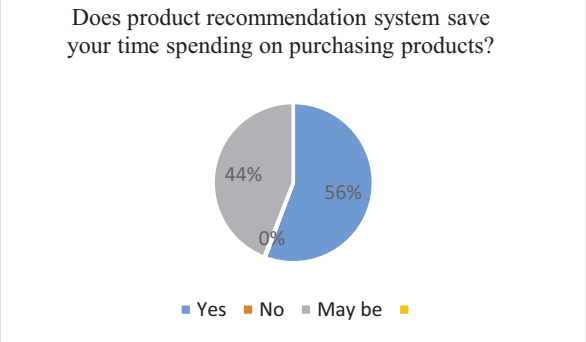

Figure:8: Pie-chart of whether product recommendation system saving time

Figure 8 shows that 56% of the people felt that product recommendation saves the time spending on purchasing products. 44% felt that product recommendation may save the time spending on purchasing products. As most of the people felt that it saves time it clearly states that there is a need for a product recommendation system in supermarkets. From the overall survey results, it was found that there is a positive impact on developing recommendation systems thus it reduces the human effort in searching. Therefore, it concludes that it is necessary to have a product recommendation system to attract both existing and new customers.

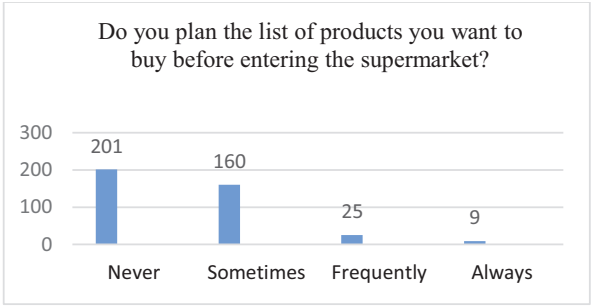

Figure:9: Bar chart of whether people plan before going to supermarket

Figure 9 shows that out 386 people, 201 people never plan the list of products before entering the supermarket, 160 people sometimes plan the list of products before entering the supermarket, 25 people frequently plan the list of products before entering the supermarket and the rest 9 people always plan the list of products before entering the supermarket. According to the results as most of the people sometimes plan the list of products before entering the supermarket it states that they don't have clear idea about what they are going to buy. Therefore, product recommendation system can able to help the people in purchasing products.

Based on the results obtained from the techniques used for product recommendation it is concluded that compared to the user-based collaborative filtering item-based collaborative filtering is better. The reason is that it is difficult to find the similar customers in a supermarket. But it is easy to find the similar items. It clearly shows that only few products can be recommended if recommendation is done using the user-based collaborative filtering. But more products can be recommended if recommendation is done using item-based collaborative filtering. The proposed system solves the problem of new customers as well as new supermarket without purchase history and product ratings. The K-Apriori algorithm effectively generates highly informative frequent item sets and association rules for the supermarket.

Based on the survey results it clearly shows that people are expecting to have a product recommendation system as they feel that it helps to purchase products and save the time spending on purchasing products. Therefore, it is necessary to have a product recommendation system in supermarket.
A well-developed recommendation system helps supermarket to improve customer's shopping experience and result in better customer retention. The system provides suggestions not only based on customer preferences but also opinions from other people.

VI. Conclusion

The recommendation system is a much sought – after powerful software solution at supermarkets for many different sensitive issues encountered by customers over choosing the products from a wide range of products with the nearest standard. The recommendation system derives much satisfaction to the customers as it reduces the time taken for the search of their desired product and helps them to choose products in the dynamic environment of a supermarket in accordance with varying preferences of customers. Recommendation system is needed in supermarkets to retain

the regular customers and to attract the new customers as well. Recommendation algorithms provide an effective form in marketing to achieve the target by furnishing the personal experience of each customer in the field of shopping. A methodology was adopted in this study for modelling and predicting the purchases at a supermarket. The main objective of this product recommendation system is to help customers to reach decisions on the purchase of preferred products through the recommendation. As far as a method of product recommendation is concerned, the purchasing history of each customer is collected and studied initially in view to increase the degree of satisfaction to potential customers. Recommendation is offered at two stages in this system. At first stage, the recommendation is given in advance in prior to choose a product while in the second stage, the recommendation is given in subsequent to the selection of a product. The proposed system is based on collaborative filtering, clustering and association rule mining. The item-to-item collaborative filtering is preferred as it provides recommendations to all customers regardless the number of purchases and in the ratings given already on priority.

Following a particular sequence will increase the accuracy of the relationship between products and the sequential processing of dataset in supermarket might yield more solid features. These would be the delightful prospects for any potential student who would prefer to select this area of study in the future to improve the existing product recommendation system.

REFERENCES

[1] K. C. A. S. A. Panayiotis Christodoulou, "A Real-Time Targeted Recommender System for Supermarkets," *19th International Conference on Enterprise Information Systems,* vol. 2, pp. 703-712, 2017.

[2] A. Demiriz, "Enhancing Product Recommender Systems on Sparse Binary Data," *Data Mining and Knowledge Discovery,* vol. 9, 2002.

[3] D. M. A. Yahya Dorostkar Navaei, "Dihedral Product Recommendation System for E-commerce Using Data Mining Applications," *International Journal of Computer & Information Technologies(IJOCIT),* vol. 3, pp. 610-631, 2015.

[4] H. Y. Duc-TrongLe, "Basket-SensitivePersonalizedItemRecommendation," *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence(IJCAI-17),* pp. 2060-2066, 2017.

[5] R. D. A. K. R. D. Debajyoti Mukhopadhyay, "A Product Recommendation System using Vector Space Model and Association Rule," *International Conference on Information Technology,* pp. 279-282, 2008.

[6] S. A. I. S. S. V. T. Andreas Kanavos, "Large Scale Product Recommendation of Supermarket Ware Based on Customer Behaviour Analysis," *Big Data and Cognitive Computing,* vol. 2, 2018.

[7] A. K. D. Loraine Charlet Annie M.C, "Market Basket Analysis for a Supermarket based on Frequent Itemset Mining," *International Journal of Computer Science Issues,* vol. 9, 2012.

[8] D. M. Ruma Dutta, "Offering A Product Recommendation System in E-commerce," 2011.

[9] B.-J. Y. J. S. S. M. K. Yong Soo Kim, "Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites," *Korea Advanced Institute of S cience and Technology,* pp. 381-393, 2005.

[10] K. C. A. S. A. P. Christodoulou, "A Real-Time Targeted Recommender System," *19th International Conference on Enterprise Information Systems,* vol. 2, pp. 703-712, 2017.